AFRL-SA-WP-TR-2016-0010

Identification of Associations Between Genetic Factors and Asthma That Are Modified by Obesity

Andrew T. DeWan, PhD
Yale University

June 2016

Final Report
for February 2013 to October 2015

Air Force Research Laboratory
711th Human Performance Wing
U.S. Air Force School of Aerospace Medicine
Aeromedical Research Department
2510 Fifth St., Bldg. 840
Wright-Patterson AFB, OH 45433-7913

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

Qualified requestors may obtain copies of this report from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-SA-WP-TR-2016-0010 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//SIGNATURE//	//SIGNATURE//
COL NICOLE ARMITAGE	DR. RICHARD A. HERSACK
Chief, En Route Care Research Division	Chair, Aeromedical Research Department

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
29 Jun 2016	Final Technical Report	February 2013 – October 2015

4. TITLE AND SUBTITLE

Identification of Associations Between Genetic Factors and Asthma That Are Modified by Obesity

5a. CONTRACT NUMBER
FA8650-13-2-6371

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Andrew T. DeWan, PhD

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Yale University
105 Wall Street
New Haven, CT 06511-6614

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
USAF School of Aerospace Medicine
Aeromedical Research Dept/FHE
2510 Fifth St., Bldg. 840
Wright-Patterson AFB, OH 45433-7913

10. SPONSORING/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-SA-WP-TR-2016-0010

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.

13. SUPPLEMENTARY NOTES
Cleared, 88PA, Case # 2016-4109, 23 Aug 2016.

14. ABSTRACT

We performed a genome-wide gene by environment (asthma) interaction analysis for the outcome of body mass index (BMI) in the Multi-Ethnic Study of Atherosclerosis (MESA) (N=2474 Caucasians, 257 asthmatics) and replicated findings in the Framingham Heart Study (FHS) offspring cohort (N=1408 Caucasians, 382 asthmatics). The replicable tagging single nucleotide polymorphism (SNP) rs2107212 was further examined in stratified analyses. Seven SNPs clustered in 17q21.2 were identified to be associated with higher BMI among asthmatics (interaction $p < 5 \times 10^{-7}$ in MESA and $p < 0.05$ in FHS). Nominally associated copy number variations interacting with BMI were associated with asthma among African American women in the Women's Health Initiative study.

15. SUBJECT TERMS
Body mass index, SNP, asthma, obesity, genome, genes

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Charles Dean
U	U	U	SAR	23	19b. TELEPHONE NUMBER *(include area code)*

This page intentionally left blank.

TABLE OF CONTENTS

Section **Page**

LIST OF FIGURES

Page

LIST OF TABLES

Page

1.0 SUMMARY

The overarching goal of this project was initially to study how obesity modifies the effects of genetic variants within previously identified genes for asthma using previously collected genotype and phenotype data from a number of existing genetic epidemiological studies. Through the course of this project, we made some interesting observations that led us to expand the scope of our search to take a genome-wide approach rather than a candidate gene approach. Further, after identifying several genome-wide significant interactions between body mass index (BMI) and single nucleotide polymorphisms (SNPs) associated with asthma, we looked at these SNPs in more detail to more fully understand the relationship between the SNPs and obesity and asthma. We made an interesting observation that the SNPs appeared to have an effect on BMI among asthmatics but not non-asthmatics. This observation was made among U.S. whites in the Multi-Ethnic Study of Atherosclerosis (MESA) dataset but not for other ethnic groups in MESA and Women's Health Initiative (WHI). None of the top SNPs overlapped among any of the ethnic groups. The results of this analysis altered the direction of this project so that we can better understand the relationship between these SNPs, BMI, and asthma.

We started by looking at BMI as an outcome among the U.S. whites from MESA. To check if interactions between significant SNPs and asthma status in MESA are replicable, the Framingham Heart Study (FHS) was used as an independent dataset to replicate the top SNPs in MESA due to the availability of subjects of similar age and ethnicity. There are three generations in the Framingham study. The second generation, the offspring sub-study, was chosen for analysis because its participants have similar age (60-70 years old) as MESA.

We performed a genome-wide gene by environment (asthma) interaction analysis for the outcome of BMI in MESA (N=2474 Caucasians, 257 asthmatics) and replicated findings in the FHS offspring cohort (N=1408 Caucasians, 382 asthmatics). The replicable tagging SNP rs2107212 was further examined in stratified analyses. Seven SNPs clustered in 17q21.2 were identified to be associated with higher BMI among asthmatics (interaction $p < 5 \times 10^{-7}$ in MESA and $p < 0.05$ in FHS). In both MESA and FHS asthmatics, subjects carrying the A allele on rs2107212 had significantly higher odds of obesity than non-carriers, which was not the case for non-asthmatics. We further examined BMI change subsequent to asthma diagnosis over a period of 26 years in FHS and demonstrated greater BMI increase among asthmatics compared to non-asthmatics. Asthmatics carrying the A allele at rs2107212 had significantly greater net BMI increase over the 26-year period compared to non-asthmatics. In this study, we found that common genetic variants on 17q21.2 are associated with post-asthma BMI increase among Caucasians.

This work is detailed in a published peer-reviewed paper: Wang L, Murk W, DeWan AT. Genome-wide gene by environment interaction analysis identifies common SNPs at 17q21.2 that are associated with increased body mass index only among asthmatics. PLoS One. 2015; 10(12):e0144114. doi: 10.1371/journal.pone.0144114. eCollection 2015.

We further explored the relationship between a different type of genetic variant, copy number variants (CNVs), which exist as deletions and duplications, asthma, and obesity. Using the WHI dataset, we analyzed the African American subjects (due to power constraints) for interactions between CNVs and BMI that were associated with asthma. The most significant deletion interacting with BMI associated with asthma is on chromosome 7 (p-value = 0.0032). The most significant duplication interacting with BMI in the outcome of asthma is on chromosome 10 (p-value = 0.0019).

2.0 INTRODUCTION

Asthma and obesity are two rapidly growing public health issues, and the comorbidity of these conditions poses an enormous burden on asthma control as well as quality of life [1]. The simultaneously increasing prevalence of both asthma and obesity suggests a potential intrinsic link between these two chronic disorders [2,3], and a recent longitudinal study identified asthma as a risk factor for subsequent obesity [4]. However, the underlying factors that contribute to this relationship have remained largely unknown.

Genetic factors play an essential role in both asthma and obesity and are believed to be predominantly responsible for the comorbidity of the two conditions [5-7]. A genomic inversion in 16p11.2 was identified to be protective against the joint occurrence of asthma and obesity in adults of European descent [8]. This provided genetic evidence of asthma-obesity co-occurrence. In terms of single nucleotide polymorphisms (SNPs), a candidate gene association study was conducted to identify shared genetic variants between childhood asthma and obesity, but no SNP was associated with both phenotypes among previously identified asthma and body mass index (BMI) genes [9]. This implied that SNPs and copy number variants (CNVs) underlying the comorbidity of asthma and obesity may exist in novel loci. A genome-wide analysis study (GWAS) was also conducted to identify genetic variants associated with BMI among 23,000 asthmatics [10]. An SNP in *DENND1B* was identified in asthmatic children in the discovery dataset but was not replicable. These two SNP studies primarily targeted childhood asthma. A potential explanation of the non-significant findings may be that the association between BMI and asthma is not as strong as in adults, since no significant BMI difference between asthmatics and non-asthmatics was observed at baseline in the GWAS study.

3.0 METHODS

3.1 Subjects and Phenotype Definition

The Multi-Ethnic Study of Atherosclerosis (MESA) and its ancillary study, MESA Air, were used for SNP analysis [11]. MESA is a population-based study focusing on characteristics and risk factors of subclinical cardiovascular disease. The study comprises 6,814 men and women aged 45-84 who were free of clinical cardiovascular diseases, recruited through six field centers across the United States. The screening exam (exam1) took place in 2000 and was followed by four examinations (exams 2-5) in 2002, 2004, 2005, and 2010, respectively. Data from exams 1-4 were available at the start of the project and were used for analysis. Height and weight were measured at every visit, and the BMI value [BMI = weight (kg)/height $(m)^2$] from the screening exam was used in the present analysis. Asthmatics were defined as those who reported doctor-diagnosed asthma in the first exam; non-asthmatics were defined as those who never reported doctor-diagnosed asthma or asthma medication use in all four exams. We included all participants in the first examination who identified themselves as Caucasian (n=2527). Subjects with missing information on asthma phenotype (n=10), inconsistent phenotype information (i.e., those who used asthma medication but did not report doctor-diagnosed asthma, n=134), or incident asthma after first examination (i.e., those who first reported asthma in exams 2-4, n=34) were excluded. MESA Air added 253 new participants, and after exclusion of subjects with missing information, we included 249 additional subjects. Of the

remaining subjects from both MESA and MESA Air (n=2598), 2588 were genotyped as part of the MESA SNP Health Association Resource (SHARe) and were included for further analysis.

Data from the Framingham Heart Study (FHS) were used for replication of the top SNPs from the discovery stage [12]. The objective of FHS is to identify common factors or characteristics that contribute to cardiovascular disease. It was established in 1948 and has developed as a prospective, community-based, three-generation study. We selected participants in the offspring cohort for analysis, since this population had a similar age distribution as the MESA population. There were eight exams from 1971 to 2005, and height and weight information was collected at each visit. BMI, calculated based on the height and weight information in the eighth examination (BMI = weight (lb)/height $(in)^2 \times 703$), was used in the present analysis. Except for the first exam, self-reported asthma and clinical diagnostic impression of asthma (CDI asthma) questions were asked at each visit. Asthmatics were defined as those who ever had self-reported asthma or wheezing in any exams. Non-asthmatics were defined as those who never had self-reported asthma and wheezing or CDI asthma. We included subjects with non-missing height and weight information in the eighth examination (n=2852). Subjects with CDI asthma but not self-reported asthma (n=13), without genotype data (n=178), or not confirmed as Caucasians (n=311) were removed, resulting in a total of 2350 subjects included in further analysis.

For the CNV analysis, we used genotype and phenotype data from the Women's Health Initiative (WHI). WHI is a long-term national health study including three clinical trials and one observational study. The goal of the observational study is to identify risk factors associated with specific health or disease outcomes. Subjects enrolled in the WHI were postmenopausal women between 50 and 79 years old, likely to live in the area for 3 years and have no condition predicting survival <3 years. The enrollment was completed in 1998 and participants were followed annually for 8 to 12 years. Genotyping was performed using Affymetrix array on African Americans (n=8218) and Hispanics (n=3482). CNVs were detected through genotype information for African Americans due to power constraints for the Hispanic subjects.

3.2 Ethical Statement

Data for both MESA, FHS, and WHI were obtained from the database of Genotypes and Phenotypes (MESA accession number: phs000209.v12.p3; FHS accession number: phs000007.v23.p8; WHI accession number: phs000200.v10.p3). This study was approved by Yale University Human Investigation Committee. Patient records/information were anonymized and de-identified prior to analysis.

3.3 Genotyping and Quality Control

In MESA, genomic DNA was isolated from peripheral blood samples and genotyped on the Affymetrix Genome-Wide Human SNP Array 6.0 chip containing 909,622 SNPs. Quality control procedures and subsequent genetic analyses were performed using PLINK 1.07 [13]. Nineteen subjects were removed because their overall call rate was <98%, and six subjects were removed because the reported sex did not match the genotyped sex as called in PLINK. SNPs were excluded if they met any one of the following criteria: 1) not mapped to an autosomal chromosome (n=37,380); 2) call rate <98% (n=21,669); 3) the genotype distributions of the SNP deviated from those expected by Hardy-Weinberg equilibrium at $p < 1\times10^{-7}$ (n=1,729);

4) monomorphic or minor allele frequency (MAF) <1% (n=126,951). A subset of linkage disequilibrium (LD)-pruned SNPs was generated using pairwise r^2>0.5 in a window of 50 SNPs and shifted by 5 SNPs at each step for genetic quality control procedures, including identity-by-descent and principal component analyses. The pairwise identity-by-descent matrix was calculated to assess cryptic relatedness. Thirty-three pairs of subjects were identified that had a pi-hat value >0.2. One subject from each pair was randomly removed. Principal component analysis was conducted using default parameters in EIGENSTRAT 3.0 [14] to address population stratification. Fifty-six outliers were identified after five iterations and excluded from further analysis. In total, 2474 subjects (n=257 asthmatics, n=2217 non-asthmatics) with 721,893 SNPs were included in the genome-wide interaction association analysis (discovery stage).

In the Framingham SHARe, a sub-study of FHS, subjects were genotyped on the Affymetrix 500K mapping array plus Affymetrix 50K supplemental array, with 500,568 SNPs in total. The same quality control procedures and criteria as used in MESA were applied to the FHS dataset. Subjects with call rate <98% (n=277) and subjects with unmatched sex check (n=8) were removed. SNPs that were not mapped to an autosomal chromosome (n=12,422), had a call rate <98% (n=79,984), failed Hardy-Weinberg equilibrium (n=1,776), or had MAF <1% or were monomorphic (n=57,689) were excluded. LD-pruned SNPs from the remaining SNPs were used for genetic quality control. One subject from each pair with pi-hat value >0.2 was randomly removed (n=563), and outliers from the principal component analysis (n=94) were removed. In total, 1408 subjects (n=382 asthmatics, n=1026 non-asthmatics) and 348,697 SNPs were included in the replication stage analysis.

In the WHI SHARe, the subjects were genotypes on the Affymetrix 6.0 mapping array, which contained probes for approximately 1,000,000 SNPs in African American (n=8218) and Hispanic (n=3482) subjects. Intensity data generated for these probes can then be utilized to perform CNV calling, detailed below.

3.4 SNP Genetic Analysis

For genetic analysis, SNPs were coded as an additive genetic model (0, 1, and 2, indicating the number of minor alleles), and BMI was treated as a quantitative trait. Asthma status was dichotomously coded. Population stratification was inspected through the genomic inflation factor (λ) in single SNP analyses, with adjustment for age, sex, asthma status, and increasing numbers of principal components. We found that the first principal component was sufficient to account for population stratification. Quantile-quantile plots were based on p-values from the genome-wide single SNP analyses. A genome-wide SNP by asthma interaction analysis was then conducted on MESA subjects, with a linear model that included main effect terms for the SNP and asthma and an interaction term for SNP × asthma. The model also included covariates to adjust for age, sex, and the first principal component. Strict Bonferroni-corrected significance was defined as $p < 6.93\times10^{-8}$ (0.05/721,893), while genome-wide suggestive significance was defined as $p < 5\times10^{-7}$ for tests of interaction terms. Three genetic regions that each had more than three genome-wide suggestive SNPs in high LD (r^2>0.8) were identified, and the SNP in each region with the smallest number of missing subjects was selected as a tag SNP for replication in FHS.

In FHS, a genome-wide single SNP analysis was first performed with a linear regression model adjusted for age, sex, asthma status, and the first principal component to inspect genomic inflation. For replication of interactions involving the tag SNPs in FHS, SNP-asthma interactions for BMI as the outcome were analyzed using linear regression models as already described. Interactions reaching nominal significance ($p<0.05$) in FHS were considered potentially replicated. The regional plot for the replicable region, 17q21.2, was made using LocusZoom [15]. The rest of the six genome-wide suggestive SNPs in 17q21.2 were then tested for interaction in FHS. SNP imputation was performed using IMPUTE2 [16] with default settings for the two un-genotyped SNPs (rs12601191 and rs16968877) in FHS. HapMap 3 haplotypes with NCBI [National Center for Biotechnology Information] build 36 (hg18) coordinates were used as the reference panel. The imputation had an overall concordance rate of 96.9%, and the info score was 0.988 for rs12601191 and 0.979 for rs16968877. All SNPs were then examined and subjected to meta-analysis, conducted using the R package "meta" [17], with both fixed effect and random effect models. The most significant interacting SNP from the meta-analysis, rs2107212, was selected for the stratification analysis.

3.5 Stratified Analysis

The association between genotypes of rs2107212 and BMI was evaluated in asthmatics and non-asthmatics in MESA and FHS subjects. A qualitative analysis using obesity (defined as BMI$\geq$30 kg/m^2) vs. non-obesity (BMI<30 kg/m^2) as the outcome was performed. A logistic regression model with adjustment for age, sex, and the first principal component was used to evaluate the odds ratio and 95% confidence interval (CI) of being obese for each additional risk allele, stratified by asthma status, in MESA and FHS and in a pooled meta-analysis. Since there was no significant heterogeneity between these studies (Cochrane's Q test $p>0.05$ and $I^2<10\%$), a fixed effect model was used in the meta-analysis. We also evaluated the association when using different BMI cut-offs of 28 kg/m^2 and 25 kg/m^2 to define obesity. In FHS, a subset of subjects who participated in exam 2 and had BMI and asthma information was extracted to examine the subjects' BMI change from 1979 (exam 2) to 2005 (exam 8). Genotype-specific mean BMI at both time points was calculated. Net BMI change over 26 years was then compared between genotypes with or without the risk allele, stratified by asthma status. Welch's t-tests were applied when comparing the difference of two groups.

3.6 CNV Detection

Two independent software packages, PennCNV [18] and Birdsuite [19], were used for CNV detection from the SNP genotyping arrays. In PennCNV, CNV calling was performed under standard protocol and adjusted for guanine-cytosine content. Sample quality controls were based on quality control metrics generated by PennCNV including Log R ratio standard deviation, B allele frequency drift, wave factor, and total number of CNVs called per person. CNVs that were less than 10 kb or contained less than 10 SNPs or resided in the centromere/telomere regions were excluded. The Birdseye module from Birdsuite was also applied to detect CNVs following standard pipeline. CNV calls with a limit of detection score > 5 were filtered in for further quality control procedures in the same way as in PennCNV. The overlapping CNV calls from both PennCNV and Birdsuite were retained for further analysis in PLINK [13].

3.7 CNV Association

First, we used two independent software packages, PennCNV and Birdsuite, to detect CNVs, respectively, in WHI and then the overlapped CNV calls were recorded for further quality control in PLINK. In PLINK, CNVs were divided into duplications and deletions and filtered out if the length was over 1000 kb or the frequency was less than 1%. Then CNVs were grouped into pools with overlapping segments. There were 681 groups of deletions and 452 groups of duplications. Genome-wide CNVs by BMI interaction analysis were performed using a logistic regression model with adjustment for age, BMI value, and the first principal component looking for associations with asthma.

4.0 RESULTS

4.1 Summary of Participants

Summary statistics of participants from MESA (n=2474) and FHS (n=1408) following quality control are shown in Table 1. The distribution of sex, age, and BMI was similar for the two datasets. Asthma prevalence differed between the two studies (10% vs. 27% in MESA and FHS, respectively). BMI was significantly higher in asthmatics compared to non-asthmatics in both studies (Welch's t-tests). As expected, the prevalence of obesity was significantly higher among asthmatics compared to non-asthmatics in both datasets (1 degree of freedom chi-square test).

Table 1. Characteristics of Subjects in MESA and FHS

Variable	MESA (n=2474)			FHS (n=1408)		
	Asthma (n=257)	Non-Asthma (n=2217)	p-value	Asthma (n=382)	Non-Asthma (n=1026)	p-value
Age (mean±SD)	62.91±10.12	61.04±9.65		68.08±8.82	68.77±8.83	
Male (%)	115 (45)	1097 (49)		189 (49)	476 (46)	
BMI (mean±SD)	29.01±6.04	27.59±6.04	3.25×10^{-4}	29.17±5.72	27.74±4.90	1.75×10^{-5}
Classification						
Underweight	3	12		4	10	
Normal weight	66	716		82	298	
Overweight	89	910		153	431	
Obesity (%)	99 (39)	579 (26)	3.37×10^{-5}	143 (37)	287 (28)	7.73×10^{-4}

SD = standard deviation.
Note: Participants who were included in the final interaction analyses are summarized. In MESA, age is based on the screening exam; in FHS, age is based on the 8th exam with 5-year intervals. Classification was based on BMI as follows: underweight (BMI<18.5); normal weight (18.5≤BMI<25); overweight (25≤BMI<30); obesity (BMI≥30).

4.2 Genome-Wide SNP by Asthma Interaction Analysis

Genome-wide SNP by asthma interaction analysis was first performed on MESA subjects. The linear regression model included main effects for SNP and asthma and an interaction term for SNP × asthma. The model was also adjusted for age, sex, and the first principal component. Overall interaction results are shown in the Manhattan plot (Figure 1). Thirty-one interacting SNPs exceeded the genome-wide suggestive threshold of $p<5\times10^{-7}$, and three genetic regions that each contained more than three genome-wide suggestive interacting SNPs in high LD with each other were selected: 7p21.3, 10q25.3, and 17q21.2. The SNPs with the least missing subjects in each region were chosen as tag SNPs for further validation in FHS. Of the three tag SNPs, one SNP (rs2107212 in 17q21.2) had a significant SNP by interaction p-value in replication subjects ($p=9.38\times10^{-3}$), while the other two SNPs did not reach the replication significance threshold of 0.05 (Table 2). To further verify the significant region, we then tested all seven MESA genome-wide suggestive SNPs in 17q21.2 in FHS. We observed that all seven SNPs had an SNP by asthma interaction replication p-value < 0.05, in the same direction as observed in MESA. The p-values for the seven SNPs ranged from 1.76×10^{-7} to 3.72×10^{-7} in MESA and 9.38×10^{-3} to 0.047 in FHS. The regional plot for SNPs located within 200 kb of the tag SNP, rs2107212, in MESA is presented (Figure 2), where a clear cluster of SNPs in strong LD is shown. SNP rs2107212 was also the most significant asthma-interacting SNP in this region following a meta-analysis of the MESA and FHS data ($p=2.51\times10^{-3}$ under a random effect model and $p=5.60\times10^{-8}$ under a fixed effect model).

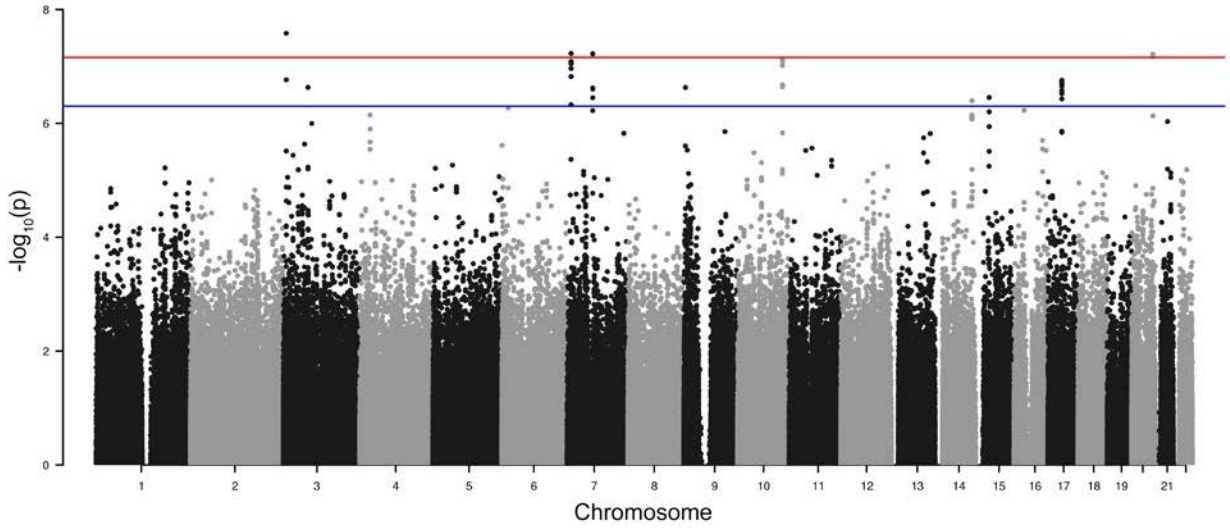

Figure 1. Manhattan plot of interaction p-values derived from the MESA genome-wide interaction analysis. *Linear regression models, including main effects for SNP and asthma and an interaction term for SNP × asthma, were used. Models were also adjusted for age, sex, and the first principal component. The red (upper) line represents the genome-wide significant p- value ($0.05/721{,}893=6.93\times10^{-8}$). The blue (lower) line represents the genome-wide suggestive p-value of 5×10^{-7}.*

Table 2. Top Interacting Regions Associated with BMI and Replication in FHS

SNP	Region	Position	Gene/Location	Dataset	β (95% CI)	p-value
rs10250689	7p21.3	10120482	*HSPA8P8/330kb up*	MESA	4.66 (2.98, 6.34)	5.91E-08
				FHS	0.56 (-0.98, 2.10)	4.75E-01
rs7904383	10q25.3	115716396	*ATRNL1/Intron*	MESA	-2.66 (-3.63, -1.70)	7.42E-08
				FHS	-0.06 (-0.94, 0.81)	8.87E-01
rs2107212	17q21.2	40949109	*KRT39/9kb down*	MESA	2.96 (1.82, 4.09)	3.72E-07
				FHS	1.49 (0.37, 2.61)	9.38E-03

β = interaction coefficient.

Note: From the MESA gene by environment interaction analysis, regions with more than three genome-wide suggestive SNPs (interacting p-value $<5\times10^{-7}$) in strong LD ($r^2>0.8$) were selected as top interacting regions. SNPs with fewer missing subjects in each region were chosen as tag SNPs to replicate in FHS. SNP position was based on the GRCh38 assembly.

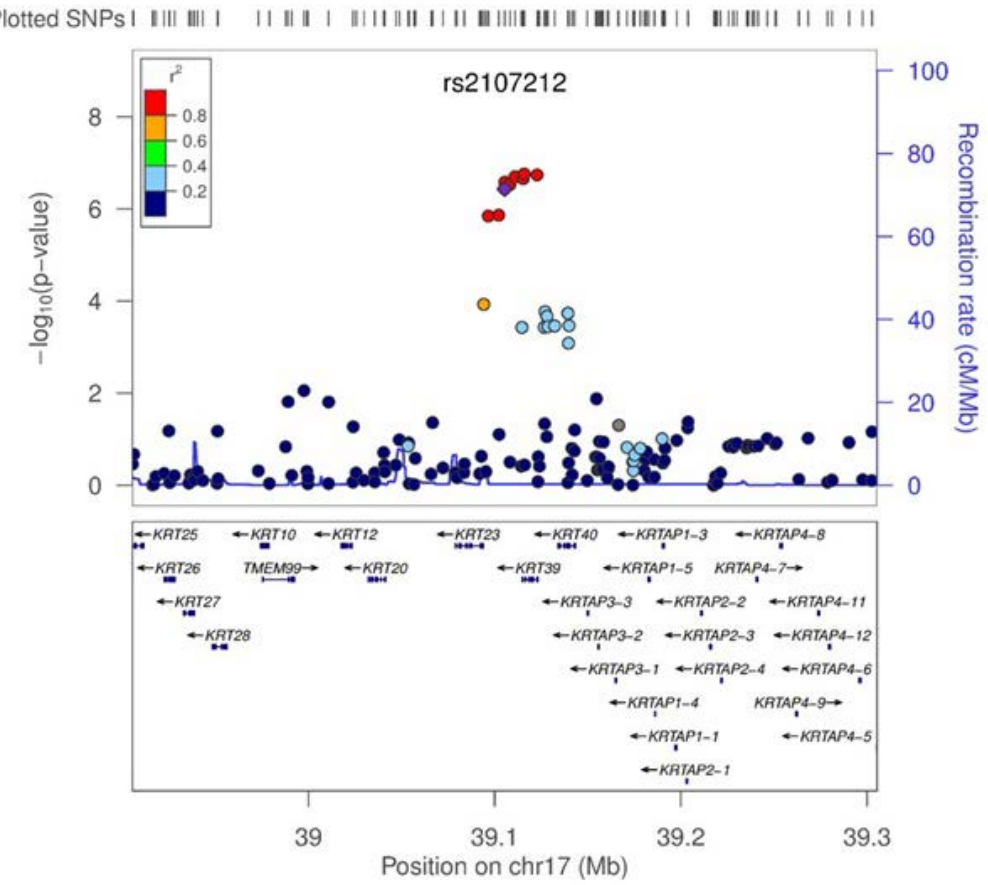

Figure 2. Regional plot showing replicable SNPs in 17q21.2. *A regional plot was generated in LocusZoom v. 1.1 for the MESA genome-wide interaction analysis results. p-values on the -log10 scale are displayed on the y axis, and chromosome positions are displayed on the x axis. This plot shows the tag SNP (rs2107212) in purple with a 200-kb flanking region on each side. The pairwise LD pattern with rs2107212 is also shown. The bottom panel shows the name, position, and transcription direction of each gene in this region.*

4.3 Association Between BMI and rs2107212 Genotypes

To further explore the nature of the interaction, we conducted a series of stratified analyses for rs2107212 among the asthmatic and non-asthmatic populations in MESA and FHS. First, we evaluated the association between rs2107212 and BMI stratified by asthma status in MESA (Figure 3A). Among asthmatics, increasing copies of the A allele at rs2107212 were associated with successively higher BMI values. Interestingly, among non-asthmatics, an opposite trend was observed, although the differences were not significant. Subjects carrying the AA or AG genotype had significantly higher BMI among asthmatics compared to non-asthmatics. In FHS, we observed a similar trend for the effect of the A allele (Figure 3B). Subjects carrying the AG genotype had a significantly greater BMI difference between asthmatics and non-asthmatics compared to those carrying the GG genotype. Subjects carrying the AA genotype had higher BMI among asthmatics compared to non-asthmatics, although the p-value did not reach nominal significance (p=0.06).

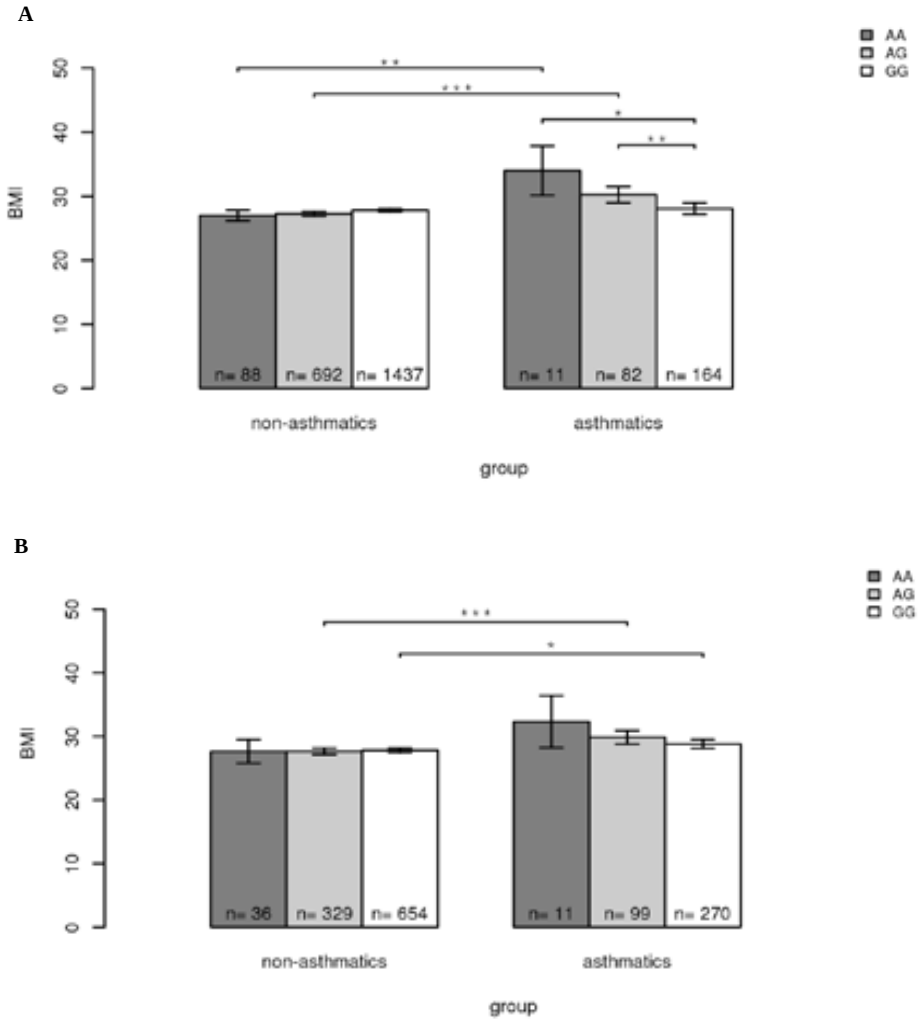

Figure 3. Association of rs2107212 genotypes and BMI, stratified by asthma status in MESA (A) and FHS (B), respectively. *BMI (kg/m^2) was compared across subjects carrying the AA, AG, or GG genotypes for rs2107212 in asthmatics and non-asthmatics. Numbers of subjects are shown in each bar. Data are represented as the mean ± standard error measurement. Significance was tested using Welch's t-test (*p<0.05, **p<0.01, and ***p<0.001).*

4.4 Evaluation of the Odds Ratio of Being Obese for the A Allele at rs2107212

We next examined and compared the odds ratio (OR) of being obese (BMI≥30 kg/m^2) among the asthmatic and non-asthmatic populations for rs2107212 in MESA and FHS separately as well as together in a meta-analysis. The forest plot showed a consistent effect of the A allele in MESA and FHS (Figure 4). The meta-analysis showed that the overall odds of being obese increased by 1.89 fold for each additional A allele in the asthmatic population. In the non-asthmatic population, the odds of being obese decreased by 0.89 fold per additional A allele; however, this was not significantly different from 1.0. To further test if the asthma-specific association results for the A allele at rs2107212 are still present with varying cut-offs for classifying obesity, we evaluated the association using different BMI cut-offs of 28 kg/m^2 and 25 kg/m^2. These results showed that different BMI cut-offs resulted in similar trends of association for rs2107212. This further demonstrated that the A allele at rs2107212 was associated with BMI increase and obesity incidence only among asthmatics.

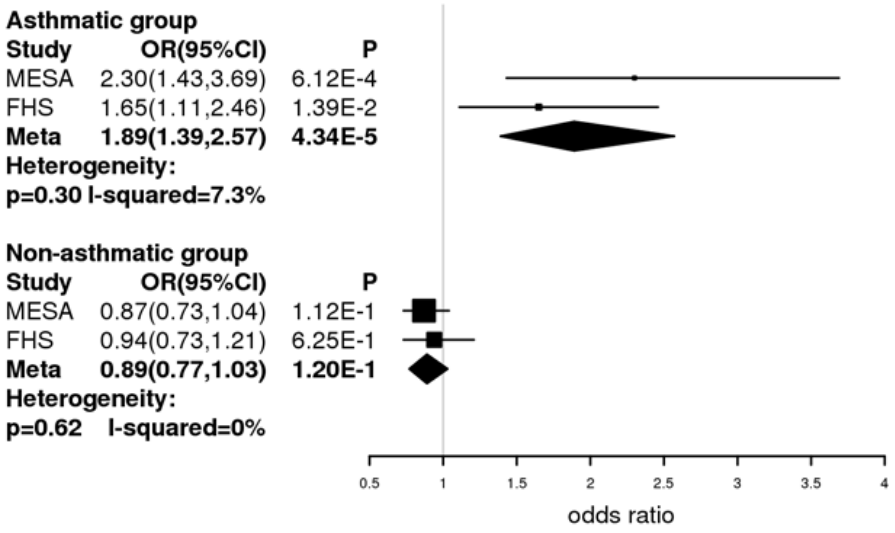

Figure 4. Forest plot showing the association of rs2107212 with obesity in MESA, FHS, and meta-analysis. *Asthma status-specific ORs of being obese for rs2107212 were calculated using logistic regression models adjusted for age, gender, and the first principal component in MESA and FHS, respectively. The minor allele A is the "risk" allele, and an additive genetic model was used. Meta-analysis was performed under a fixed effect model since the heterogeneity tests were not significant. Boxes indicate group-specific OR point estimates, and lines indicate the respective 95% CI. Diamonds indicate meta-analysis OR and 95% CI.*

4.5 BMI Change Over a 26-Year Period After Asthma Diagnosis in FHS

In FHS, 990 subjects had BMI and asthma data for both exams 2 (1979) and 8 (2005), of which 983 subjects had genotype information for rs2107212. Mean BMI at these two times, stratified by asthma status and rs2107212 genotype, is shown in Figure 5A. BMI among asthmatics was significantly higher than in non-asthmatics in 1979 (p=0.0102), and the difference was more significant in 2005 ($p=6.84\times10^{-4}$). During this 26-year period, in asthmatics, subjects carrying the AA or AG genotype increased their BMI more than subjects carrying the GG genotype; however, this was not true among non-asthmatics (Figure 5B). During this time period, asthmatics carrying the AA or AG genotypes had a mean BMI change of 3.79, while

subjects carrying the GG genotype had a mean BMI change of 3.24. Among non-asthmatics, subjects carrying the AA or AG genotypes had a mean BMI change of 2.31, while subjects carrying the GG genotype had a mean BMI change of 2.71. This result further indicates that the asthmatic population in general gained more weight than the non-asthmatic population and that the A allele at rs2107212 increased the BMI change by 60% among asthmatics compared to non-asthmatics (p=0.0308).

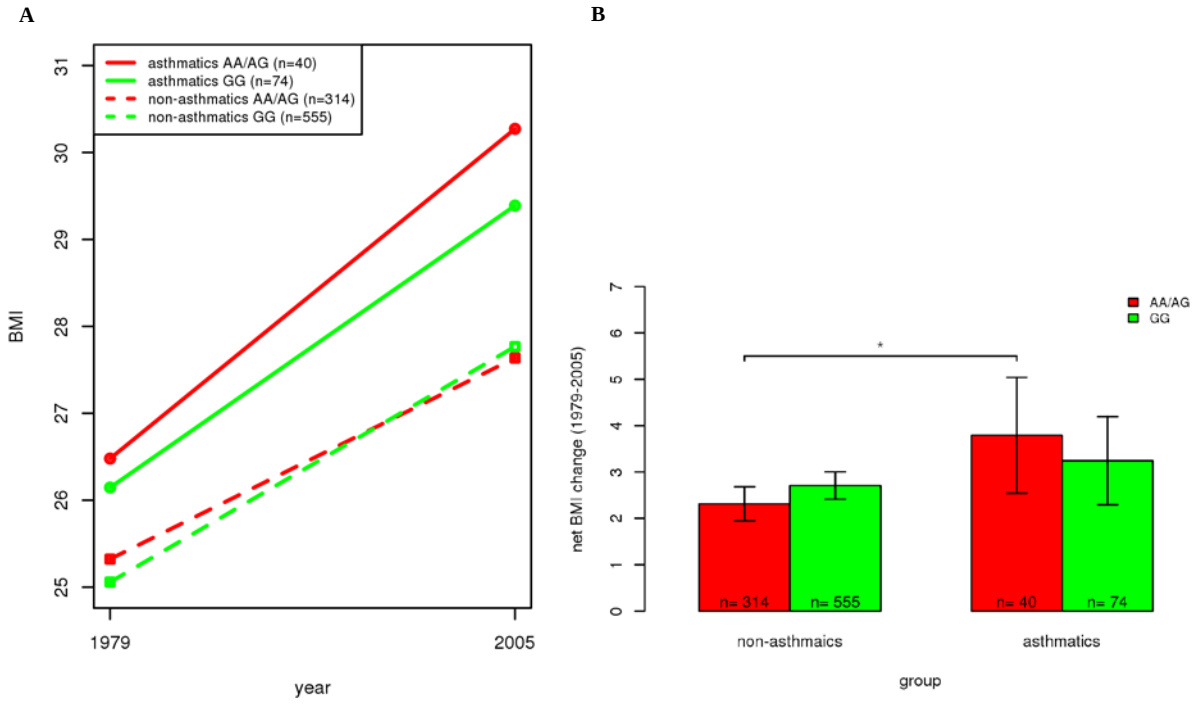

Figure 5. Association of rs2107212 genotypes and BMI change by asthma status in FHS. *(A) For a subset of FHS subjects who had BMI and asthma information at both exam 2 (1979) and exam 8 (2005) (n=983), the means of BMI at both time points are shown according to the rs2107212 genotypes AA/AG and GG, stratified by asthma status. (B) Net BMI change (BMI value in 2005 – BMI value in 1979) with rs2107212 genotypes AA/AG and GG stratified by asthma status is shown. The number of subjects is shown in each bar. Data are represented as the mean ± standard error measurement. *p<0.05 under Welch's t-test.*

4.6 CNV Results

We have identified top CNVs for African Americans in the WHI dataset. The most significant deletion interacting with BMI for asthma is on chromosome 7, from 117927028 base pairs (bp) to 117944067 bp (p-value = 0.0032; Table 3). The most significant duplication region that is interacting with BMI in the outcome of asthma is in chromosome 10, from 46496693 bp to 46499590 bp (p-value = 0.0019; Table 4).

Table 3. Top 5 Deletions in Interaction Analysis

Chromosome	Starting bp	Ending bp	OR	95% CI	p-value
7	117927028	117944067	1.07	1.02, 1.12	0.0032
1	110044476	110044476	1.18	1.04, 1.33	0.0080
9	43666998	43666998	1.08	1.02, 1.15	0.0119
2	35437965	35459760	1.07	1.01, 1.13	0.0130
1	110038455	110038470	1.15	1.03, 1.28	0.0146

Table 4. Top 5 Duplications in Interaction Analysis

Chromosome	Starting bp	Ending bp	OR	95% CI	p-value
10	46496693	46499590	1.11	1.04, 1.19	0.0019
10	46501701	46501866	1.11	1.04, 1.19	0.0024
10	46504791	46504871	1.11	1.04, 1.18	0.0030
10	46506308	46506308	1.11	1.04, 1.18	0.0030
10	46510869	46510869	1.10	1.03, 1.17	0.0039

5.0 DISCUSSION

Many correlated disorders have shared genetic backgrounds, and the genetic effects on one of the phenotypes might be modified by the other phenotype. This provides a novel model supporting the integration of gene by environment interaction terms (i.e., where the "environment" is another disease or phenotype) in a GWAS framework to explore comorbid susceptibility genes. These genetic factors could be overlooked when only SNP main effects are examined.

The comorbidity of asthma and obesity is a growing medical problem. While previous studies have focused on the unidirectional relationship of obesity preceding asthma onset, post-asthma BMI increase is a neglected but potentially important phenomenon, particularly since asthma is an early onset disease. Here, we used two independent datasets of older adults and hypothesized that asthma impacted BMI through genetic factors. Therefore, we conducted an SNP by asthma interaction analysis to identify SNPs that are associated with BMI and modified by asthma status. SNPs in 17q21.2 were verified to be associated with higher BMI and higher risk of obesity only when asthma is present.

We further took advantage of the longitudinal FHS dataset to obtain BMI measured closest to asthma onset and evaluated BMI change after asthma diagnosis from 1979 to 2005. Figure 5A reflects the trend of BMI change over this 26-year period. In 1979, the asthmatic population already had significantly higher BMI than non-asthmatics. The BMI differential further increased in the intervening 26 years and resulted in a much greater BMI differential among asthmatics than non-asthmatics. When genotype was taken into consideration, asthmatic subjects carrying the minor allele of rs2107212 gained more weight than subjects not carrying this allele, while this was not the case among non-asthmatics. Since BMI in 1979 was likely measured after but closer to asthma onset than BMI in 2005, we can infer that the BMI difference between asthmatics and non-asthmatics before asthma onset was less than in 1979 and that actual BMI increase after asthma may be greater. In Figure 5B, we further examined the net BMI change over a 26-year period, where rs2107212 functioned as a risk locus for BMI increase

only when asthma was present. Here, we identified a genetic locus that is associated with post-asthma BMI increase and demonstrated that asthma could influence weight gain over a long period.

Chromosome 17q21 has been repeatedly reported to be associated with asthma [20-22]. Although the region covering the ORMDL3, GSDMA, and GSDMB genes is the most replicated one, the keratin (KRT) cluster has also been shown to be associated with asthma in a previous GWAS study [23]. KRT proteins are key structural components in epithelial cells and are essential for tissue function. A study identified KRT18 as a bronchial epithelial autoantigen that is associated with adulthood nonallergic asthma [24]. KRT18 and some other keratin genes have also been identified to be associated with obesity-caused fatty liver [25,26]. Hair follicles have been demonstrated to be associated with Toluene diisocyanate-induced asthma [27]. Non-functional mutations in filaggrin, a protein binding to and condensing the keratin cytoskeleton, are risk factors for asthma, atopic eczema, and allergies [28]. The identified SNP rs2107212 is located in the intergenic region flanked by keratin genes KRT39 and KRT23 (9 kb upstream of KRT39 and 11 kb downstream of KRT23). KRT39 is a type I keratin gene with expression late in the differentiation of hair [29]. KRT23 is a type I epithelial keratin gene that has been identified to interact with 14-3-3 proteins to modulate key cellular processes in a SMAD4-dependent manner [30]. In addition, KRT23 has been reported to be associated with several cancers, such as colon cancer, pancreatic cancer, and hepatocellular carcinoma [31-33]. This study links the keratin family genes with asthma and obesity phenotypes. However, the underlying biological mechanisms of rs2107212 in asthma-dependent BMI increase need to be further explored.

This study is potentially confounded by reverse causality from obesity to asthma. Longitudinal studies have observed an increased OR or relative risk of incident asthma among the obese population [34,35]. We cannot exclude the possibility of a causal effect of obesity on asthma, since BMI before asthma development was not available in MESA and the Framingham offspring study. However, to attempt to further assess SNP effects on BMI increase after asthma diagnosis, we used a subset of subjects from the Framingham offspring cohort who had both asthma status and BMI values in 1979 and 2005. Over a period of 26 years, the net BMI change was significantly higher in asthmatics than non-asthmatics. This result demonstrates that SNP rs2107212 is involved in the effect of asthma on BMI. However, from the data we have, we cannot determine if rs2107212 is also involved in the reverse relationship, BMI on asthma.

Another limitation in this study is that MESA and FHS did not use the same asthma definitions. MESA asked about doctor-diagnosed asthma, while FHS asked about wheezing and asthma in the same question and a doctor diagnosis was not required. This could explain the asthma prevalence difference (10% in MESA and 27% in FHS). A relatively looser asthma definition in FHS may have potentially diluted any signals and may have underestimated some of the interactions, which we believed was the primary reason for lack of replication of other top interacting regions. However, the consistency in direction of effect between MESA and FHS for the tag SNPs strongly argues in favor of these being true positive interactions.

After adjusting for the number of SNPs tested in the genome-wide screen, this study had over 80% power to detect a β of 5 when the SNP MAF is >0.1 and a β of 4 when the SNP MAF is >0.2. While this is a modest level of power, this dataset allowed us to identify a novel and replicable gene × environment interaction.

The results of the CNV by BMI analysis in African Americans in the WHI study were not genome-wide significant when looking for associations with asthma. However, the top deletion and duplication were nominally $p<0.05$ and may be worth further exploration in the future in other datasets containing African Americans and/or other ethnic groups.

6.0 CONCLUSIONS

In conclusion, this is the first attempt at a genome-wide interaction analysis to detect genetic risk factors associated with BMI modified by asthma. SNPs in 17q21.2 were identified as risk loci for BMI increase only among asthmatics. This finding will help elucidate pathways involved in the comorbidity of asthma and obesity. Further, we identified some candidate CNVs that interact with BMI that are associated with asthma. Additional studies are needed to further verify and emphasize the relationship between asthma, obesity, and genetic markers and examine the underlying mechanisms.

7.0 REFERENCES

1. Ford ES. The epidemiology of obesity and asthma. J Allergy Clin Immunol. 2005; 115(5), 897-909; quiz 910.
2. Stukus DR. Obesity and asthma: the chicken or the egg? J Allergy Clin Immunol. 2015; 135(4):894-895.
3. Kim SH, Sutherland ER, Gelfand EW. Is there a link between obesity and asthma? Allergy Asthma Immunol Res. 2014; 6(3):189-195.
4. Egan KB, Ettinger AS, DeWan AT, Holford TR, Holmen TL, Bracken MB. Longitudinal associations between asthma and general and abdominal weight status among Norwegian adolescents and young adults: the HUNT Study. Pediatr Obes. 2015; 10(5):345-352.
5. Meyers DA, Bleecker ER, Holloway JW, Holgate ST. Asthma genetics and personalised medicine. Lancet Respir Med. 2014; 2(5):405-415.
6. Winter Y, Sankowski R, Back T. Genetic determinants of obesity and related vascular diseases. Vitam Horm. 2013; 91:29-48.
7. Hallstrand TS, Fischer ME, Wurfel MM, Afari N, Buchwald D, Goldberg J. Genetic pleiotropy between asthma and obesity in a community-based sample of twins. J Allergy Clin Immunol. 2005; 116(6):1235-1241.
8. González JR, Cáceres A, Esko T, Cuscó I, Puig M, et al. A common 16p11.2 inversion underlies the joint susceptibility to asthma and obesity. Am J Hum Genet. 2014; 94(3):361-372.
9. Melén E, Himes BE, Brehm JM, Boutaoui N, Klanderman BJ, et al. Analyses of shared genetic factors between asthma and obesity in children. J Allergy Clin Immunol. 2010; 126(3):631-637.e1-e8.
10. Melén E, Granell R, Kogevinas M, Strachan D, Gonzalez JR, et al. Genome-wide association study of body mass index in 23 000 individuals with and without asthma. Clin Exp Allergy. 2013; 43(4):463-474.
11. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, et al. Multi-Ethnic Study of Atherosclerosis: objectives and design. Am J Epidemiol. 2002; 156(9):871-881.
12. Dawber TR, Meadors GF, Moore FE Jr. Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health. 1951; 41(3):279-281.

13. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81(3):559-575.
14. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38(8):904-909.
15. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010; 26(18):2336-2337.
16. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3 (Bethesda). 2011; 1(6):457-470.
17. Schwarzer G. General package for meta-analysis, version 4.4-1. 2016. [Accessed 1 Aug 2016]. Available from https://cran.r-project.org/web/packages/meta/meta.pdf.
18. Wang K, Li M, Hadley D, Liu R, Glessner J, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007; 17(11):1665-1674.
19. Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat Genet. 2008; 40(10):1253-1260.
20. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med. 2010; 363(13):1211-1221.
21. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. Nature. 2007; 448(7152):470-473.
22. Ono JG, Worgall TS, Worgall S. 17q21 locus and ORMDL3: an increased risk for childhood asthma. Pediatr Res. 2014; 75(1-2):165-170.
23. Laprise C. The Saguenay-Lac-Saint-Jean asthma familial collection: the genetics of asthma in a young founder population. Genes Immun. 2014; 15(4):247-255.
24. Nahm DH, Lee YE, Yim EJ, Park HS, Yim H, et al. Identification of cytokeratin 18 as a bronchial epithelial autoantigen associated with nonallergic asthma. Am J Respir Crit Care Med. 2002; 165(11):1536-1539.
25. Park JE, Kim HT, Lee S, Lee YS, Choi UK, et al. Differential expression of intermediate filaments in the process of developing hepatic steatosis. Proteomics. 2011; 11(14):2777-2789.
26. Watanabe T, Takemura M, Saito K, Ito H, Hattori T, et al. [Clinical significance of the measurement of serum cytokeratin-18 in patients with non-alcoholic steatohepatitis] [Article in Japanese]. Rinsho Byori. 2013; 61(1):19-24.
27. Nayak AP, Hettick JM, Siegel PD, Anderson SE, Long CM, et al. Toluene diisocyanate (TDI) disposition and co-localization of immune cells in hair follicles. Toxicol Sci. 2014; 140(2):327-337.
28. Irvine AD, McLean WH, Leung DY. Filaggrin mutations associated with skin and allergic diseases. N Engl J Med. 2011; 365(14):1315-1327.
29. Langbein L, Rogers MA, Praetzel-Wunder S, Böckler D, Schirmacher P, Schweizer J. Novel type I hair keratins K39 and K40 are the last to be expressed in differentiation of the hair: completion of the human hair keratin catalog. J Invest Dermatol. 2007; 127(6):1532-1535.

30. Liffers ST, Maghnouj A, Munding JB, Jackstadt R, Herbrand U, et al. Keratin 23, a novel DPC4/Smad4 target gene which binds 14-3-3ε. BMC Cancer. 2011; 11:137.
31. Zhang JS, Wang L, Huang H, Nelson M, Smith DI. Keratin 23 (K23), a novel acidic keratin, is highly induced by histone deacetylase inhibitors during differentiation of pancreatic cancer cells. Genes Chromosomes Cancer. 2001; 30(2):123-135.
32. Wang K, Xu X, Nie Y, Dai L, Wang P, Zhang J. Identification of tumor-associated antigens by using SEREX in hepatocellular carcinoma. Cancer Lett. 2009; 281(2):144-150.
33. Birkenkamp-Demtroder K, Christensen LL, Olesen SH, Frederiksen CM, Laiho P, et al. Gene expression in colorectal cancer. Cancer Res. 2002; 62(15):4352-4363.
34. Nystad W, Meyer HE, Nafstad P, Tverdal A, Engeland A. Body mass index in relation to adult asthma among 135,000 Norwegian men and women. Am J Epidemiol. 2004; 160(10):969-976.
35. von Kries R, Hermann M, Grunert VP, von Mutius E. Is obesity a risk factor for childhood asthma? Allergy. 2001; 56(4):318-322.

LIST OF ABBREVIATIONS AND ACRONYMS

β	interaction coefficient
BMI	body mass index
bp	base pairs
CDI	clinical diagnostic impression
CI	confidence interval
CNV	copy number variant
FHS	Framingham Heart Study
GWAS	genome-wide analysis study
KRT	keratin
LD	linkage disequilibrium
MAF	minor allele frequency
MESA	Multi-Ethnic Study of Atherosclerosis
OR	odds ratio
SHARe	SNP Health Association Resource
SNP	single nucleotide polymorphism
WHI	Women's Health Initiative